nature portfolio

Corresponding author(s): Clifford A. Meyer

Last updated by author(s): Mar 22, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	N/A
Data analysis	MetaTiME is publicly available at https://github.com/yi-zhang/MetaTiME. A tutorial is available at https://github.com/yi-zhang/MetaTiME/blob/main/docs/notebooks/metatime_annotator.ipynb. MetaTiME is also available at Zhang Y et al., "MetaTiME Integrates Single-cell Gene Expression to Unveil the Meta-components of the Tumor Immune Microenvironment", MetaTiME, https://doi.org/10.5281/zenodo.7734062, 2023. Software used: MetaTiME==1.3.0, pandas==1.1.5, scanpy==1.8.2, anndata==0.8.0, matplotlib==3.5.1, adjustText==0.7.3; leidenalg==0.8.3, harmonypy==0.0.9, sklearn==0.22, GSEApy==0.10.4, scipy==1.6.3, clusterProfiler==4.0.5, singleR==1.8.0,Lisa(https://github.com/liulab-dfci/lisa2), scsim(https://github.com/dylkot/scsim), Seurat==v4.0.5, umap-learn==0.5.3.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our <u>policy</u>

The processed datasets used in this study are available in Zenodo under accession code 7410180: https://doi.org/10.5281/zenodo.7410180. The pretrained meta-components for tumor microenvironment in the Github repository: https://github.com/yi-zhang/MetaTiME. The single-cell RNA-seq data used in this study are available in the TISCH database (http://tisch1.comp-genomics.org). The list of public datasets used in this study is available in Supplementary Data 1 and also from GEO under accession code GSE154763 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE154763) and Single Cell Portal under accession code SCP1288 (https://singlecell.broadinstitute.org/single_cell/study/SCP1288/tumor-and-immune-reprogramming-during-immunotherapy-in-advanced-renal-cell-carcinoma). The gene list analytical data used in this study are available in AnimalTFDB v3.0 (http://bioinfo.life.hust.edu.cn/AnimalTFDB/#!/), TCGA (https://portal.gdc.cancer.gov), and Azimuth (https://azimuth.hubmapconsortium.org). The ChIP-seq data used in this study are available from GEO (https://www.ncbi.nlm.nih.gov/geo/) under accession IDs GSM604651, GSM1637309, GSM1607526, GSM2661793, GSM2735378, GSM2871705, GSM1637306, GSM1637307, and from ENCODE (https://www.encodeproject.org/) under accession ID ENCSR919OXR.

Human research participants

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

[x] Life sciences [ ] Behavioural & social sciences [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The number of datasets are determined using the largest possible tumor scRNA dataset from TISCH.
Data exclusions	Quality control on single-cell data was done to filter low-quality cells.
Replication	Replication of results are made by simulating knockout datasets.
Randomization	N/A since no experimental groups are designed.
Blinding	Random knockout of dataset was performed to obtain results.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	Antibodies
☒	Eukaryotic cell lines
☒	Palaeontology and archaeology
☒	Animals and other organisms
☒	Clinical data
☒	Dual use research of concern

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging